AD-A045 159 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1

AN ALMOST SURE REPRESENTATION FOR INTERMEDIATE ORDER STATISTICS--ETC(U)

SEP 77 V WATTS N00014-75-C-0809

UNCLASSIFIED MIMEO SER-1136 NL

1 OF 1

AD

A045159

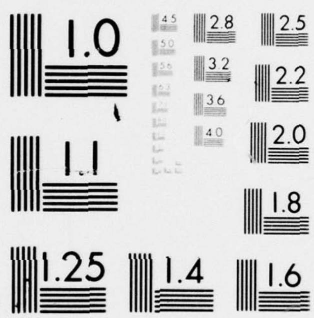
END

DATE

FILMED

11-77

DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

An Almost Sure Representation for Intermediate Order Statistics: The Finite Endpoint Case

Vernon Watts

Institute of Statistics Mimeo Series No. 1136
September, 1977

Department of Statistics
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 An almost sure representation for intermediate order statistics: The finite endpoint case		5. TYPE OF REPORT & PERIOD COVERED 9 TECHNICAL rept.
		6. PERFORMING ORG. REPORT NUMBER 14 Mimeo Ser[redacted]-1136
7. AUTHOR(s) 10 Vernon Watts		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-75-C-0809
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 12 13 p.
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Arlington, VA 22217		12. REPORT DATE 11 Sep[redacted] 1977
		13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

intermediate order statistics, almost sure representation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let $\{X_n\}_{n\geq 1}$ be independent and identically distributed and let $X_{k_n}^{(n)}$ denote the k_n-th smallest order statistic for $X_1,\ldots,X_n$, where $k_n \to \infty$ but $k_n/n \to 0$. A representation for $X_{k_n}^{(n)}$ in terms of the empirical distribution function is developed, and other limiting properties are discussed. It is assumed that there is a real number x_o such that $F(x_o) = 0$, where F is the marginal distribution of X_1.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

182 850 mt

An Almost Sure Representation for Intermediate Order Statistics: The Finite Endpoint Case

by Vernon Watts*

1. Introduction.

Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables on (Ω, F, P) with marginal distribution function $F(x) = P(X_1 \leq x)$. Let $\{k_n\}_{n\geq 1}$ be a rank sequence of integers, that is, $1 \leq k_n \leq n$ for each n, and denote by $X^{(n)}_{k_n}$ the k_n-th smallest order statistic of the sample $X_1, \ldots, X_n$. Suppose that $F(x_\lambda) = \lambda$, for a given λ in the open unit interval $(0, 1)$, and in a neighborhood $(x_\lambda - \delta, x_\lambda + \delta)$ the second derivative F'' exists and is bounded, and assume that $F'(x_\lambda) > 0$. Let $F_n(x, \omega) = n^{-1} \sum_{i=1}^{n} I_{[X_i \leq x]}$ be the empirical distribution function of $X_1, \ldots, X_n$. Then for the sequence of central order statistics $\{X^{(n)}_{k_n}\}_{n\geq 1}$, where $k_n = n\lambda + o(n^{\frac{1}{2}} \log n)$, Bahadur (1966) showed that

$$X^{(n)}_{k_n}(\omega) = x_\lambda + \frac{k_n - nF_n(x_\lambda, \omega)}{nF'(x_\lambda)} + R_n(\omega), \tag{1.1}$$

where $R_n(\omega) = \mathcal{O}(n^{-3/4} \log n)$ as $n \to \infty$ with probability one. The precise order of the remainder term R_n was found by Kiefer (1967). From the representation (1.1) may be obtained important limiting properties of central order statistics, including asymptotic normality and the law of the iterated logarithm. This representation has been extended to central order statistics from dependent sequences $\{X_n\}$ by Sen (1968, 1972), and under fewer assumptions upon the marginal d.f. F a weaker convergence of R_n was obtained by Ghosh (1971).

In this paper we establish an analogous representation for intermediate order statistics $\{X^{(n)}_{k_n}\}$, that is, when the rank sequence $\{k_n\}$ satisfies $k_n \to \infty$ but $\frac{k_n}{n} \to 0$. ($\{k_n\}$ is then called an intermediate rank sequence.) For simplicity we deal only with the case of i.i.d. sequences $\{X_n\}$, although extensions

*Research sponsored by the Office of Naval Research N00014-75-C-0809.

to dependent or non-identically distributed sequences are possible. Our assumptions upon the marginal d.f. F parallel those of Bahadur and involve the behavior of F near its left endpoint x_o, defined by $x_o = \inf\{x: F(x) > 0\}$, which here we assume to be finite. Additionally, we obtain from our representation the asymptotic normality property and a law of the iterated logarithm for intermediate order statistics.

2. Statement and proof of the main result.

Let F be a d.f. with a finite left endpoint x_o such that $F(x_o) = 0$ (that is, F has no probability mass at x_o). Suppose that inside an interval $(x_o, x_o + \delta)$ F is twice differentiable with F" bounded, and that $\lim_{x \downarrow x} F'(x)$ exists and is positive. Let $\{k_n\}$ be an intermediate rank sequence; we impose the mild restriction that $\frac{k_n}{\log n} \to \infty$. We then determine x_n by the relation $F(x_n) = \frac{k_n}{n}$; the value x_n is well-defined and unique for all sufficiently large n, by the assumptions upon F.

Again denote the empirical d.f. of $X_1, \ldots, X_n$ by $F_n(x,\omega)$. In this section we prove the following result.

Theorem 1. Suppose that $\{X_n\}$ is an i.i.d. sequence with marginal d.f. F satisfying the above left endpoint conditions, and let $\{k_n\}$ be an intermediate rank sequence for which $\log n = o(k_n)$ as $n \to \infty$. Then

$$X_{k_n}^{(n)}(\omega) = x_n + \frac{k_n - nF_n(x_n,\omega)}{nF'(x_n)} + R_n(\omega), \tag{2.1}$$

where

$$R_n(\omega) = O(n^{-1} k_n^{1/4} \log^{3/4} n) \text{ as } n \to \infty$$

with probability one.

Throughout we suppose that $\{X_n\}$ and $\{k_n\}$ are given sequences of r.v.'s and integers, respectively, for which the hypotheses of Theorem 1 are satisfied. Define a sequence of positive numbers

$$a_n = n^{-1} k_n^{1/2} (1 + \varepsilon)^{1/2} \log^{1/2} n$$

where $\varepsilon > 0$, and a sequence of integers $\{b_n\}$ satisfying

$$b_n \sim \left(\frac{k_n}{\log n}\right)^{1/4} \text{ as } n \to \infty.$$

For sufficiently large n let

$$G_n(x, \omega) = F_n(x, \omega) - F(x) - F_n(x_n, \omega) + \frac{k_n}{n}$$

and

$$H_n(\omega) = \sup_{x_n - a_n \le x \le x_n + a_n} |G_n(x, \omega)|.$$

The following two lemmas constitute the major portion of the proof of Theorem 1.

Lemma 2.1. With probability one,

$$H_n(\omega) = O(n^{-1} k_n^{1/4} \log^{3/4} n) \text{ as } n \to \infty.$$

<u>Proof</u>. Let N_1 be an integer such that x_n is uniquely defined for $n \geq N_1$. For $n \geq N_1$ and for any integer r, let $\eta_{r,n} = x_n + a_n b_n^{-1} r$. By the monotonicity of F and F_n we have for $x \in [\eta_{r,n}, \eta_{r+1,n}]$,

$$F_n(\eta_{r,n}, \omega) - F(\eta_{r+1,n}) \leq F_n(x, \omega) - F(x) \leq F_n(\eta_{r+1,n}, \omega) - F(\eta_{r,n}).$$

Thus

$$H_n(\omega) \leq \max_{-b_n \leq r \leq b_n} \{|G_n(\eta_{r,n}, \omega)|\}$$

$$+ \max_{-b_n \leq r \leq b_n - 1} \{F(\eta_{r+1,n}) - F(\eta_{r,n})\}$$

$$(2.2) \qquad = K_n(\omega) + \alpha_n,$$

say. Since $\eta_{r+1,n} - \eta_{r,n} = a_n b_n^{-1}$, $\eta_{b_n,n} = x_n + a_n \to x_o$, and by the assumptions upon F in the fixed interval $(x_o, x_o + \delta)$, we have that

$$(2.3) \qquad \alpha_n = O(n^{-1} k_n^{1/4} \log^{3/4} n).$$

We will use the following inequality of Bernstein (see Bahadur (1966)): Let $B(n, p)$ be a binomial r.v. with parameters n and p, for some integer $n \geq 1$ and for $0 \leq p \leq 1$. Then for every $t > 0$,

$$(2.4) \qquad P(|B(n, p) - np| \geq t) \leq 2 \exp(-h),$$

where $h = h(n, p, t)$ is given by

$$h = \frac{t^2}{2\{np(1 - p) + \frac{t}{3} \max\{p, 1 - p\}\}} .$$

Let $C_1 > \lim_{x \downarrow x_0} F'(x) > 0$. We may choose $N_2 \geq N_1$ such that $F(x_n + a_n) - F(x_n) < C_1 a_n$ and $F(x_n) - F(x_n - a_n) < C_1 a_n$, for all $n \geq N_2$. Now for any $n \geq N_1$ and any integer r,

$$G_n(\eta_{r,n}) = \begin{cases} n^{-1} \sum_{i=1}^{n} I_{[x_n < X_i \leq \eta_{r,n}]} - p_{r,n}, & \text{if } r \geq 0 \\ -n^{-1} \sum_{i=1}^{n} I_{[\eta_{r,n} < X_i \leq x_n]} + p_{r,n}, & \text{if } r \leq -1, \end{cases}$$

where $p_{r,n} = |F(\eta_{r,n}) - \frac{k_n}{n}|$. Letting $t_n = C_2\, n^{-1} k_n^{1/4} \log^{3/4} n$ for some $C_2 > 0$, we therefore have by (2.4),

$$P(|G_n(\eta_{r,n})| \geq t_n) \leq 2 \exp(-h_{r,n}),$$

where

$$h_{r,n} = \frac{n^2 t_n^2}{2\{np_{r,n}(1 - p_{r,n}) + \frac{nt_n}{3} \max\{p_{r,n}, 1 - p_{r,n}\}\}}$$

$$\geq \frac{nt_n^2}{2\, p_{r,n} + t_n} .$$

Since $n \geq N_2$ and $|r| \leq b_n$ imply $p_{r,n} < C_1 a_n$, we have

$$P(|G_n(\eta_{r,n})| \geq t_n) \leq 2 \exp(-\delta_n)$$

for $n \geq N_2$ and $|r| \leq b_n$, where

$$\delta_n = \frac{nt_n^{\;2}}{2\, C_1\, a_n + k_n} ,$$

which does not depend on r. Therefore

$$P(K_n \geq t_n) \leq \sum_{r=-b_n}^{b_n} P(|G_n(\eta_{r,n})| \geq t_n)$$

$$\leq 4\, b_n \exp\,(-\delta_n) \;.$$

An easy calculation shows that $\delta_n \geq \dfrac{C_2^{\;2}}{2\, C_1\,(1+\varepsilon)^{1/2} + 1} \log n$ for $n \geq N_3$, say,

so that for given C_1 and ε, choosing C_2 sufficiently large such that

$\dfrac{C_2^{\;2}}{2\, C_1\,(1+\varepsilon)^{1/2} + 1} > \dfrac{5}{4}$, we obtain

$$\sum_{n \geq N_1} P(K_n \geq t_n) < \infty \;.$$

Thus by the Borel-Cantelli lemma,

$$P(K_n \geq t_n \;\text{ infinitely often}) = 0,$$

and since $t_n = C_2\, n^{-1} k_n^{1/4} \log^{3/4} n$, the statement of the lemma follows from (2.2) and (2.3).

Lemma 2.2. With probability one,

$$x_n - a_n \leq X_{k_n}^{(n)}\,(\omega) \leq x_n + a_n$$

for all large n (that is, for $n \geq$ some $N(\omega)$). Note that $x_o < x_n - a_n$ for all large n, by the choice of a_n, the conditions on F, and the restriction $\log n = o(k_n)$.

Proof. First we have

$$P(X_{k_n}^{(n)} < x_n - a_n) \leq P\left(\sum_{i=1}^{n} I_{[X_i \leq x_n - a_n]} \geq k_n\right)$$

$$= P\left(\sum_{i=1}^{n} I_{[X_i \leq x_n - a_n]} - np_n \geq k_n - np_n\right)$$

$$\leq P\left(\sum_{i=1}^{n} I_{[X_i \leq x_n - a_n]} - np_n > C_3 n a_n\right)$$

where $p_n = F(x_n - a_n)$, for some $C_3 > 0$, and for all large n, since $a_n = o(\frac{k_n}{n})$ and since F' exists and is bounded away from zero in some fixed interval. Then by (2.4),

$$P(X_{k_n}^{(n)} < x_n - a_n) \leq 2 \exp\left(\frac{-C_3^2 n^2 a_n^2}{2np_n + C_3 n a_n}\right)$$

$$\leq 2 \exp\left(\frac{-C_4 n^2 a_n^2}{k_n}\right)$$

for some $C_4 > 0$ (not depending on ε) and for large n, since $p_n < \frac{k_n}{n}$. But $k_n^{-1} n^2 a_n^2 = (1 + \varepsilon) \log n$, so that if $\varepsilon > 0$ is chosen sufficiently large, we obtain

$$\sum_{n \geq N_1} P(X_{k_n}^{(n)} < x_n - a_n) < \infty . \tag{2.5}$$

In a similar way we have

$$P(X_{k_n}^{(n)} > x_n + a_n) \leq P\left(\sum_{i=1}^{n} I_{[X_i > x_n + a_n]} - np_n > C \; na_n\right)$$

for large n, where now $p_n = 1 - F(x_n + a_n)$. Then the relations $F(x_n + a_n) < \frac{k_n}{n} + C_1 a_n$ and $a_n = o(\frac{k_n}{n})$ along with (2.4) lead to

$$P(X_{k_n}^{(n)} > x_n + a_n) \leq 2 \exp \left(\frac{-C_5 n^2 a_n^2}{k_n}\right)$$

for some $C_5 > 0$ and for large n, so if ε is sufficiently large,

$$\sum_{n \geq N_1} P(X_{k_n}^{(n)} > x_n + a_n) < \infty . \tag{2.6}$$

Combining (2.5) and (2.6) and applying the Borel-Cantelli lemma completes the proof.

Proof of Theorem 1: From Lemmas 2.1 and 2.2 it follows that with probability one,

$$|F_n(X_{k_n}^{(n)}(\omega), \omega) - F(X_{k_n}^{(n)}(\omega)) - F_n(x_n, \omega) + \frac{k_n}{n}| = O(n^{-1} k_n^{1/4} \log^{3/4} n).$$

Since F has a bounded second derivative in $(x_o, x_o + \delta)$ and since $a_n = o(1)$, Lemma 2.2 gives

$$F(X_{k_n}^{(n)}(\omega)) = \frac{k_n}{n} + (X_{k_n}^{(n)}(\omega) - x_n)F'(x_n) + O(n^{-2} k_n \log n)$$

w.p.1. Then since we may assume (w.p.1.) that $F_n(X_{k_n}^{(n)}(\omega),\omega) = \frac{k_n}{n}$, noting that $F(x_o) = 0$, and from the assumption that $F'(x_n)$ is bounded away from zero for large n, the theorem follows.

3. Further Limiting Properties.

If $\{k_n\}$ satisfies the slightly more severe restriction $\frac{k_n}{\log^3 n} \to \infty$, then we may verify using the representation (2.1) that under the assumptions of the last section, the intermediate order statistic $X_{k_n}^{(n)}$ is asymptotically normal with mean x_n and variance $\frac{k_n}{(nF'(x_n))^2}$. For this we merely apply the ordinary central limit theorem for triangular arrays to $X_{n,i}$ defined by

$$X_{n,i} = k_n^{-\frac{1}{2}} (I_{[X_i \leq x_n]} - \frac{k_n}{n}).$$

This result has been previously obtained by Cheng (1965), without assuming the existence of the second derivative F".

Additionally, we may use (2.1) to establish as a new result a law of the iterated logarithm for $X_{k_n}^{(n)}$. This has been found explicitly by Kiefer (1972, Theorems 5 and 6) for intermediate order statistics from i.i.d. sequences of r.v.'s uniformly distributed on the unit interval, under the restrictions that $k_n \to \infty$ and $\frac{k_n}{n} \to 0$ monotonically and $\frac{k_n}{\log \log n} \to \infty$. Suppose the d.f. F is everywhere continuous, so that each $F(X_i)$ has the uniform (0,1) distribution. Then translating Kiefer's Theorem 5 according to the relation

$$F_n(x_n, \omega) = n^{-1} \sum_{i=1}^{n} I_{[F(X_i) \leq \frac{k_n}{n}]}$$

for sufficiently large n, we obtain from (2.1) the following result.

Theorem 2. Let $\{X_n\}$ be i.i.d. with continuous marginal d.f. F satisfying the conditions stated in Section 2. If the intermediate rank sequence $\{k_n\}$ satisfies $k_n \uparrow \infty$, and $\frac{k_n}{n} \downarrow 0$, and

$$\frac{k_n (\log \log n)^2}{\log^3 n} \to \infty ,$$

then

$$P(\overline{\lim_n} \pm \frac{nF'(x_n)(X_{k_n}^{(n)} - x_n)}{(2 k_n \log \log n)^{1/2}} = 1) = 1.$$

Finally we mention that an asymptotic approximation to the second moment of the remainder term defined by (2.1) may be derived using a procedure similar to that of Duttweiler (1973) to estimate the quantity ER_n^2 for R_n given in the Bahadur representation (1.1). To do this it is somewhat more convenient to deal with the slightly different intermediate order statistic representation

$$X_{k_n}^{(n)}(\omega) = x_n' + \frac{\frac{k_n}{n+1} - F_n(x_n',\omega)}{F'(x_n')} + R_n'(\omega), \tag{3.1}$$

where x_n' satisfies $F(x_n') = \frac{k_n}{n+1}$ (for large n). It is easily seen from the proofs in Section 2 that also $R_n'(\omega) = O(n^{-1} k_n^{1/4} \log^{3/4} n)$ as $n \to \infty$ with probability one. Then under the restriction that $\frac{k_n}{n^\theta} \to \infty$ for some $\theta > 0$, we have that for R_n' defined by (3.1),

$$E(R_n')^2 \sim (\tfrac{2}{\pi})^{1/2} \frac{k_n^{1/2}}{(nF'(x_n'))^2} \quad \text{as} \quad n \to \infty .$$

The details of the calculation of this approximation are provided by Watts (1977).

References

Bahadur, R. R. (1966). A note on quantiles in large samples. Ann. Math. Statist. 37, 577-580.

Cheng, B. (1965). The limiting distributions of order statistics. Chinese Math. 6, 84-104.

Duttweiler, D. L. (1973). The mean-square error of Bahadur's order-statistic approximation. Ann. Statist. 1, 446-453.

Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist. 42, 1957-1961.

Kiefer, J. (1967). On Bahadur's representation of sample quantiles. Ann. Math. Statist. 38, 1323-1342.

Kiefer, J. (1972). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. Proc. Sixth Berkeley Symp. Math. Statist. Probabil. 1, 227-244.

Sen, P. K. (1968). Asymptotic normality of sample quantiles for m-dependent processes. Ann. Math. Statist. 39, 1724-1730.

Sen, P. K.(1972). On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. J. Mult. Analysis 2, 77-95.

Watts, J. H. V. (1977). Limit theorems and representations for order statistics from dependent sequences. Ph.D. dissertation, Univ. of N.C.